DOCUMENT RESUME

ED 358 737 FL 021 316

AUTHOR Berne, Jane E.
TITLE The Role of Text Type, Assessment Task, and Target Language Experience in L2 Listening Comprehension Assessment.
PUB DATE 93
NOTE 45p.; Paper presented at the Annual Meetings of the American Association for Applied Linguistics and the American Association of Teachers of Spanish and Portuguese (74th, Cancun, Mexico, August 9-13, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Evaluation Methods; *Language Role; *Language Tests; *Listening Comprehension; *Listening Comprehension Tests; Second Language Learning; *Spanish; Test Construction; *Test Format; Testing

ABSTRACT

A study investigated the comparability of second language (L2) listening comprehension tests by examining the effects of varying text type and assessment task on student performance. Student target language experience was an additional variable considered. Subjects were 107 beginning and 64 advanced-intermediate college-level Spanish second language learners. Stimulus materials were a lecture text and an interview text on the same topic and with comparable content. The assessment tasks were a multiple-choice task, an open-ended task, and a cloze task, each consisting of 10 items. Tasks were randomly assigned. Relationships between text type, assessment task type, score, and language experience were analyzed. Results indicate that, contrary to previous research findings, text type alone was not a significant factor in L2 listening comprehension performance. However, assessment task was a significant factor in performance, with multiple-choice tasks producing the highest scores. Students with more target language experience had higher scores on all evaluations, but language experience did not change the relative performance on different task types. With one exception, the three variables did not interact. The single interaction occurred between test type and assessment task on items testing comprehension details. Several recommendations for instruction and testing resulting from this study are offered. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED358737

The Role of Text Type, Assessment Task, and Target Language Experience in L2 Listening Comprehension Assessment

Jane E. Berne, Ph. D.
University of North Dakota

Address: Department of Modern and Classical Languages
University of North Dakota
Box 8198
Grand Forks, ND 58202 U.S.A.

Phone: (701) 777-4653

e-mail: BERNE@NDSUVM1.bitnet

FL021316

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jane E. Berne
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

The Role of Text Type, Assessment Task, and Target Language Experience in L2 Listening Comprehension Assessment

This study addresses the issue of the comparability of second language (L2) listening comprehension assessment instruments by examining the effects of varying text type and assessment task on the listening comprehension performance of two levels of learners of Spanish as a Foreign Language. Results revealed that assessment task and target language experience are significant factors in overall L2 listening comprehension performance but that text type is not. Further analysis revealed that text type may be a significant factor in the comprehension of details, but only on certain types of assessment tasks. Recommendations regarding L2 listening comprehension instruments are offered.

The Role of Text Type, Assessment Task, and Target Language Experience in L2 Listening Comprehension Assessment [1,2]

Introduction

Over the years, researchers such as Postovsky (1981) and Krashen (1982) have pointed out the important role listening comprehension skills play in the development of other second language (L2) skills. As a result, L2 instructors are placing greater emphasis on the development of listening comprehension skills. This increased concern for the development of listening skills has led L2 instructors to pay greater attention to the assessment of listening comprehension performance.

Previous research has shown that L2 listening comprehension performance may be affected by variables such as age (Seright, 1985), gender (Bacon, 1991), background knowledge (Markham & Latham, 1987), increased exposure to authentic language (Herron & Seay, 1991), and different types of speech modifications and simplifications (Chaudron, 1983; Long, 1985). In light of this data, it is logical to conclude that L2 listening comprehension performance may also be affected by varying the stimulus materials and/or assessment tasks used to assess it.

Stimulus materials refer to the texts to which subjects are exposed in a testing situation while assessment tasks refer to the tasks or tests used to measure comprehension. Among the most commonly used stimulus materials in listening comprehension assessment are: short lectures; stories; descriptions; dialogues;

interviews; news or weather reports; and oral directions or instructions. Examples of verbal assessment tasks commonly used in listening comprehension assessment include: true/false, multiple-choice, and open-ended questions; cloze passages; recall protocols; and summaries. Non-verbal tasks such as drawing a picture or diagram, indicating a route on a map, and acting out a response have also been used to assess listening comprehension.

Each of these stimulus materials and assessment tasks differs from the others in a number of ways which may affect listening comprehension performance. However, to date there has been very little research which evaluates and compares the effects of different stimulus materials and assessment tasks on L2 listening comprehension performance. In an effort to address this question, this study compared how varying one aspect of the stimulus materials, namely text type, affects L2 listening comprehension performance across different types of assessment tasks and across different levels of target language experience. It was hoped that the insights provided by this study might make it possible to suggest recommendations with regard to the preparation and/or scoring of L2 listening comprehension assessment instruments.

Review of the Literature

Stimulus Materials

Results of a number of studies have revealed that stimulus materials are a significant factor in L2 comprehension performance (Shohamy, 1984; Steffenson and Joag-Dev, 1984; Markham and Latham, 1987). There are several ways in which stumulus materials may differ, and thus affect subjects' ability to demonstrate their comprehension of the texts in question. Stimulus materials vary in terms of their content and also in terms of their lexical and grammatical complexity. A third dimension along which stimulus materials may vary is text type, that is, their format and the genre to which they belong. Text type has been employed as a variable in L2 reading comprehension research (Shohamy, 1984; Carrell 1984); however, such comparisons have rarely been undertaken in L2 language listening comprehension research.

One study of L2 listening comprehension which employed text type as a variable is reported in Shohamy and Inbar (1991). In their study, Shohamy and Inbar compared the listening comprehension performance of Hebrew speaking learners of English as a Foreign Language across three different text types: a news broadcast, a lecturette, and a consultative dialogue. Despite identical factual and lexical content, the three texts differed in several ways. The news broadcast was characterized by complex sentences, formal language, and a lack of redundancies, repetitions, and pauses. In addition, the

speaker did not interact with an audience and the information presented was assumed to be novel information rather than shared knowledge. The lecturette was characterized by a mixture of complex and simple sentences, more familiar language, and contained many redundancies, repetitions, and pauses. In this text, the speaker interacted with an audience and the information presented was assumed to be shared knowledge. The consultative dialogue was characterized by simple sentences, colloquial language, and many redundancies, repetitions, and pauses. There was constant interaction between a presumed expert and an addressee and there was also a high degree of assumed shared knowledge.

Results of this study indicated that subjects who listened to either the lecturette or the consultative dialogue received significantly higher scores on an open-ended comprehension task than those who listened to the news broadcast. Shohamy and Inbar attributed these findings to the fact that the dense and concise nature of the news broadcast text may have impeded comprehension while the redundancies, repetitions, pauses, and monitoring of information flow that characterized the lecturette and consultative dialogue texts may have better enabled the subjects to activate relevant strategies, thereby facilitating comprehension. In addition, Shohamy and Inbar argued that from a pragmatic standpoint, the familiarity of the language employed and the amount of interaction with the audience further facilitated comprehension of the lecturette and consultative dialogue texts vis a vis the news broadcast text.

The present study sought in part to confirm Shohamy and

Inbar's findings by investigating the relationship between text type and L2 listening comprehension performance in a different language learning context, namely with English speaking learners of Spanish as Foreign Language. In the present study, it was decided to employ only two text types: a lecture and an interview. The interview text was judged to be comparable to the consultative dialogue employed in the Shohamy and Inbar study in that there was interaction between an expert and an addressee, there was a certain amount of shared knowledge, and there were repetitions and redundancies built in to the text. The lecture text was judged to be comparable to the news broadcast employed in the Shohamy and Inbar study in that there was no interaction with the audience, little or no shared knowledge, and there were few redundancies, repetitions, and pauses built in to the text. As in the Shohamy and Inbar study, factual and lexical content was kept constant across both texts to control for affects of background knowledge.

Beyond attempting to confirm Shohamy and Inbar's findings, this study also sought to address additional questions regarding the role of text type in L2 listening comprehension performance. Specifically, this study sought to examine the relationship between text type and assessment task in determining L2 listening comprehension performance and whether this relationship changes as a function of target language experience. Such questions were not addressed in the Shohamy and Inbar study since only one type of assessment task was employed and the subjects who participated in their study had similar levels of target language experience. The

variables of assessment task and target language experience are discussed in greater detail in the following sections.

Assessment Task

Previous research has found that assessment task is a significant factor in L2 comprehension performance (Shohamy, 1984, Lee, 1987; Rubin and Roberts, 1987, Wolf, 1991). As with stimulus materials, there are several ways in which assessment tasks may differ and thus affect subjects' ability to demonstrate their comprehension of the texts in question. Assessment tasks may vary according to a number of factors including: language of assessment, type of response required, and skills required to complete the task. The effects of different assessment tasks have been investigated in L2 reading comprehension research (Shohamy, 1984; Lee, 1987; and Wolf 1991). However, with exception of the study reported in Rubin and Roberts (1987), the relative effects of different assessment tasks on L2 listening comprehension performance have not been investigated extensively.

In their study, Rubin and Roberts employed both multiple-choice and open-ended tasks. Specifically, they compared subjects' performance on open-ended and multiple-choice versions of the Communicative Competency Assessment Instrument (CCAI). They judged the open-ended version of the CCAI to be superior to the multiple-choice version because subjects' performance on the open-ended version of the CCAI correlated more highly with their

performance on other tests of listening comprehension than did subjects' performance on the multiple-choice version of the CCAI.

However, this conclusion must be viewed with caution. Subjects' superior performance on the open-ended version of the CCAI may be due to several factors: the questions comprising the open-ended version of the CCAI were presented orally while the questions comprising the multiple-choice version were presented in written form; subjects taking the open-ended version had more time to answer the questions; subjects taking the open-ended version were able to ask the rater to repeat or rephrase a question if they did not understand it; and finally, subjects taking the open-ended version were not required to read in the second language.

In order to investigate the possible effects of assessment task on L2 listening comprehension performance, the present study compared subjects' performance on three types of assessment tasks: a multiple-choice task, an open-ended task, and a cloze passage. These three assessment tasks were chosen because they represent three of the most common types of assessment tasks used in both comprehension research and standardized comprehension testing. The limitations mentioned with regard to the Rubin and Roberts study were avoided by presenting all items on each of the three assessment tasks in written form and by using the subjects' native language as the language of assessment.

Target Language Experience

The final variable employed in this study was subjects' target language experience. For purposes of this study, target language experience was defined as the amount of formal classroom study of the target language that a particular second or foreign language learner has had. Much of the L2 comprehension literature reveals that subjects with higher levels of target language experience significantly outperform subjects with lower levels of target language experience (e.g., Hudson, 1982; Shohamy, 1984; Lee, 1987; VanPatten, 1990; and Wolf, 1991). Target language experience was included in this study mainly in order to investigate possible qualitative differences in the patterns of performance of different levels of learners. In other words, this study sought to address the question of whether the listening comprehension performance of different levels of learners varies across different text types and different assessment tasks. To address this question, two levels of learners of Spanish as a Foreign Language participated in this study: beginning level learners and advanced-intermediate level learners. These two levels of target language experience were chosen because they represent the bulk of learners enrolled in foreign language programs across the United States.

Research Questions and Hypotheses

The specific research questions addressed by the study were the following:

1) Does foreign language learners' listening comprehension performance vary as a function of text type?

2) Does foreign language learners' listening comprehension performance vary as a function of assessment task?

3) Does foreign language learners' listening comprehension performance vary as a function of target language experience?

4) Does foreign language learners' listening comprehension performance vary as a function of text type, assessment task, and target language experience acting in combination?

With respect to the first research question, it was hypothesized that text type would prove to be a significant factor in foreign language learners' listening comprehension performance. Based on the results of Shohamy and Inbar (1991), it was further hypothesized that subjects listening to the interview text would score significantly higher than those listening to the lecture text. Regarding the second research question, it was hypothesized that assessment task would prove to be a significant factor in foreign

language learners' listening comprehension performance. Based on the results of Shohamy (1984) and Wolf (1991), it was further hypothesized that subjects completing the multiple-choice task would score significantly higher than subjects completing either the open-ended or the cloze task.

Concerning the third research question, it was hypothesized that target language experience would prove to be a significant factor in foreign language learners' listening comprehension performance. Based on the results of numerous studies, it was further hypothesized that the advanced-intermediate level subjects would score significantly higher than the beginning level subjects. Finally, with respect to the fourth research question, the null hypothesis was adopted. That is, it was hypothesized that the three variables would not interact to any significant degree. This hypothesis was adopted as a result of the equivocal nature of previous research regarding the relationship between these three variables.

Subjects

In order to investigate the effects of target language experience, the performance of 107 beginning and 64 advanced-intermediate learners of Spanish was compared. All subjects were native speakers of English studying Spanish as a Foreign Language at a large midwestern university. The beginning level subjects were enrolled in a fourth-semester basic Spanish course. They were classified as beginning level learners because, in accordance with

Lee's (1988) timeline, they had had less than 300 hours of target language instruction. The advanced-intermediate subjects consisted of students in their third and fourth years of Spanish at the university level who had completed at least four courses beyond the basic language curriculum.

Materials [3]

The stimulus materials employed in this study consisted of a lecture text and an interview text. Both texts were prepared by the researcher and dealt with the topic of using videotaped resumés as a means of applying for a job. The lecture text consisted of an 833 word lecturette with a running time of six minutes and 26 seconds. The interview text consisted of a 1,040 word formal interview with a running time of seven minutes and seven seconds. In order to ensure that the information in both texts correlated as closely as possible, the interview text was developed principally by paraphrasing or reproducing the content of the lecture text. At the same time, an effort was made to differentiate the interview text from the lecture text by including elements of negotiated discourse such as repetitions and restatements. Once the texts were prepared, they were evaluated by native speakers of Spanish to ensure that the language employed in both texts was characteristic of oral rather than written Spanish. Both texts were then recorded on videotape by native speakers of Spanish.

The assessment tasks employed in this study consisted of a multiple-choice task, an open-ended task, and a cloze task. Each assessment task consisted of 10 items. The multiple-choice consisted of 10 incomplete statements in English with a set of three options for completing each statement. The open-ended task consisted of the same 10 incomplete statements used in the multiple-choice task but with the options for completing the statement deleted. Subjects were directed to complete the statements in English. The cloze passage consisted of a 498 word summary in English of the information in both texts. Ten phrases or clauses were deleted using a rational deletion procedure. Subjects were directed to fill in the deleted information in English. The phrases and clauses which were deleted corresponded to the responses on the multiple-choice and open-ended tasks. This ensured that the items on all three assessment tasks were parallel.

Data Collection Procedures

At the time of testing, packets containing an assessment task, and a background questionnaire were randomly distributed among the subjects so that each of the three assessment tasks was completed by approximately one-third of the subjects. Subjects were given two minutes to examine the test items or the cloze passage. This was done as a pre-listening activity in an effort to help subjects focus their listening. At the end of the allotted two minutes, subjects were instructed to close their packets and the testing commenced.

Before the videotape of the appropriate text was played, subjects were told that they were going to watch a lecture or interview similar to those that they would see as part of a televised news or information program. They were instructed to listen carefully and to try to remember as much of the information as possible. The videotape was played once and while the videotape was playing, subjects were not allowed to take notes. These procedures were adopted as they reflect authentic listening behavior when watching television. That is, when watching television, the viewer has only one opportunity to hear the information and he or she does not typically take notes. When the videotape finished playing, subjects were instructed to turn once again to the assessment task and complete it. No time limit was imposed for the completion of the assessment task.

Scoring Procedures

Subjects' scores on the different assessment tasks were determined by counting the number of correct responses out of a possible 10 correct responses. Thus, subjects' scores could range from zero to 10. On the multiple-choice task, a response was judged to be correct if the subject had indicated the correct response. For the open-ended and cloze tasks, an acceptable word criteria was adopted rather than an exact word criteria. That is, a response was judged to be correct if it was deemed to be an acceptable response, regardless of whether or not it contained the exact wording of the

researcher-prepared response.

The data were first scored by the researcher. Lists of acceptable and unacceptable responses were developed by the researcher and then reviewed by an independent rater. Any discrepancies were discussed by the two raters and final lists of acceptable and unacceptable responses were drawn up by the researcher. Then, in order to ensure the reliability of the scoring, approximately 10 percent of the data were scored by a second independent rater. Since scoring was relatively objective, it was decided that a small sampling of the data would be sufficient to determine the reliability of the scoring. Interrelater reliability was determined to be .99.

Results

Once the data were scored, they were submitted to a three-way Analysis of Variance (ANOVA) with a 2 X 3 X 2 factorial design in order to determine the effects of text type, assessment task, and target language experience on subjects' listening comprehension performance. For purposes of this study, alpha was set at the .05 level. The subjects' raw scores on the assessment task they completed constituted the dependent variable while text type, assessment task, and target language experience constituted the independent variables. All three independent variables were treated as between-group variables. The results of the ANOVA, shown in Table 1, indicated main effects for assessment task (F (2,159) =

12.498, p = .0001) and for target language experience (F (1,159) = 56.294, p = .0001). There was no main effect for text type and there were no interactions.

Insert Table 1

The main effects for assessment task and target language experience were submitted to post-hoc Scheffé's tests in order to determine the source of the significant difference. Results of the Scheffé's test for assessment task, shown in Table 2, revealed that scores on the multiple-choice task were significantly higher than scores on either the open-ended (p = .0001) or the cloze task (p = .0044), whereas scores on the open-ended and cloze tasks did not differ significantly (p = .0954). The results of the Scheffé's test for target language experience, reported in Table 3, revealed that the advanced-intermediate level subjects scored significantly higher than the beginning level subjects (p = .0001).

Insert Table 2

Insert Table 3

Given that the hypothesized main effect for text type was not obtained, additional analyses were conducted in an effort to uncover any significant patterns with regard to text type that were not

revealed by the ANOVA. The first additional analysis consisted of examining subjects' performance at the item level in order to determine if subjects' performance on specific items might have obscured any significant patterns with regard to text type. The item analysis consisted of calculating a series of item means and item-to-total correlations. Item means represent the proportion of subjects who responded correctly on each item. Item-to-total correlations measure the degree to which subjects' performance on particular items reflects their performance on the task as a whole.

As a first step in the item analysis, item-to-total correlations for each of the 10 items across the entire data set were computed. For purposes of this analysis, both Pearson and Spearman rank order correlation coeffients were computed. The resulting sets of rank order correlation coefficients are displayed in Table 4. Both sets of rank order correlation coefficients for the entire data set were consistent across all 10 items, suggesting that the test items were uniformly difficult for all subjects.

Insert Table 4

This conclusion left open the possibility that the performance of subjects in particular experimental cells might be obscuring significant patterns with regard to text type. A second set of item-to-total correlations were computed for all 10 items across 12 different subgroups of the data set, representing the 12 experimental cells. An examination of this data revealed no consistent patterns

with regard to the performance of the different groups of subjects on eight of the 10 items. On two of the 10 items, Items 6 and 8, means of zero were obtained for four of the 12 groups of subjects. A mean of zero for a particular cell on either or both of these two items indicated that no subject in that cell responded correctly on either one or both of the items.

The high incidence of zero means on Items 6 and 8 suggested the possibility that the extreme difficulty experienced by some groups of subjects on these items might have obscured siginificant patterns with regard to text type. It was decided to reanalyze the data excluding the two difficult items to see if any significant patterns with regard to text type emerged. This follow-up analysis on the adjusted scores consisted of a second three-way ANOVA with a 2 X 3 X 2 factorial design. The independent variables for this second ANOVA were the same as for the previous ANOVA; however, in this analysis, the dependent variable consisted of the subjects' adjusted scores rather than their overall scores. The adjusted scores were computed by subtracting the subjects' scores on the two difficult items from their overall scores.

The results of the second ANOVA, shown in Table 5, mirror the results of the first ANOVA. There were main effects for assessment task ($F(2,159) = 9.974$, $p = .0001$) and target language experience ($F(1,159) = 51.033$, $p = .0001$), but there was no main effect for text type nor were there any interactions. Post-hoc Scheffé's tests for the main effects for assessment task and target language experience further confirmed that subjects' performance on the adjusted task

was similar to their performance on the overall task. Scores on the multiple-choice task were significantly higher than scores on either the open-ended task ($p = .0001$) or the cloze task ($p = .0167$) whereas scores on the open-ended and cloze tasks did not differ significantly ($p = .1179$). The advanced-intermediate level subjects scored significantly higher than the beginning level subjects ($p = .0001$).

Insert Table 5

Since none of the previous analyses revealed any significant patterns with regard to text type, it was decided to conduct further analyses using substantively interesting subsets of items rather than individual items or the task as a whole as the focus of the analyses. Given that main ideas and details represent information from different levels of the text, it was thought that they might be sensitive to differences in text type. Shohamy and Inbar (1991) investigated the relationship between text type and question type and found that the interactions between text type and local questions (testing comprehension of details) were higher than the interactions between text type and global questions (testing comprehension of main ideas). They also found that for both local and global question types, the pattern of the interactions was similar. The interactions between the questions and the consultative dialogue were the highest followed by the interactions between the questions and the lecturette which were followed by the interactions between the

questions and the news broadcast.

In the present study, to determine if there were any differences in subjects' comprehension of main ideas and details, their performance on items testing comprehension of main ideas and items testing comprehension of details was analyzed via two additional three-way ANOVAs with a 2 X 3 X 2 factorial design. The independent variables for both ANOVAs were the same as for the previous ANOVAs; however, for the first of these additional ANOVAs, the dependent variable consisted of the subjects' scores on items testing comprehension of main ideas while for the second, the dependent variable consisted of the subjects' scores on items testing comprehension of details.

In order to determine which items tested comprehension of main ideas and which tested comprehension of details, copies of the lecture text were distributed to eight experienced instructors of Spanish who were then asked to mark which sentences or phrases represented main ideas. Material left unmarked was assumed to represent details. Responses were tallied by the researcher. To make the final determination of whether a particular item tested comprehension of main ideas or details, a simple majority of the instructors had to agree. Five items were determined to be testing comprehension of main ideas and five items were determined to be testing comprehension of details.

Results of the ANOVA for main idea scores, displayed in Table 6, were similar to the results of the previous ANOVAs. There were main effects for assessment task ($F(2,159) = 17.056$, $p = .0001$) and

target language experience ($F(1,159) = 58.349$, $p = .0001$), but there was no main effect for text type and there were no interactions. Post-hoc Sheffé's tests for both assessment task and target language experience futher demonstrated that subjects' performance on items testing comprehension of main ideas was similar to their performance on the both the task as a whole and the adjusted task. Main idea scores on the multiple-choice task were significantly higher than main idea scores on either the open-ended task ($p = .0001$) or the cloze task ($p = .0001$) whereas main idea scores on the open-ended and cloze tasks did not differ significantly ($p = .9074$). The advanced-intermediate level subjects scored significantly higher than beginning level subjects ($p = .0001$).

Insert Table 6

Regarding the ANOVA for detail scores, shown in Table 7, results revealed main effects for assessment task ($F(2,159) = 4.598$, $p = .0114$) and target language experience ($F(1,159) = 19.1869$, $p = .0001$). Once again, there was no main effect for text type, but there was an interaction between text type and assessment task ($F(2, 159) = 3.936$, $p = .0215$). No other interactions were obtained.

Insert Table 7

Due to the presence of an interaction, post-hoc Scheffés tests was deemed to be inappropriate. In order to investigate the source

of the main effects and the interaction, means were submitted to a series of post-hoc contrasts. Results of the post-hoc contrasts for assessment task, shown in Table 8, revealed that detail scores on the open-ended task were significantly lower than detail scores on either the the multiple-choice ($F = 7.917$, $p = .0055$) or the cloze task ($F = 6.000$, $p = .0154$) whereas detail scores on the multiple-choice and cloze tasks did not differ significantly ($F = .187$, $p = .6659$). The post-hoc constrast for target language experience, shown in Table 9, revealed that the advanced-intermediate level subjects scored significantly higher than the beginning level subjects ($F = 19.869$, $p = .0001$).

Insert Table 8

Table 9

In order to investigate the interaction between text type and assessment task, a series of five post-hoc contrasts was run. Results of these contrasts, displayed in Table 10, revealed that subjects completing the multiple-choice task were able to recall more details from the interview text than they were from the lecture text ($F = 6.808$, $p = .0099$). No other contrasts reached significance.

Insert Table 10

Discussion and Conclusion

Based on the results of this study, several conclusions can be drawn. Regarding the first research question, the results of this study revealed, in contrast to the results obtained by Shohamy and Inbar (1991), that text type alone is not a significant factor in L2 listening comprehension performance. In four separate analyses, no main effect for text type was obtained. The only effect for text type occurred with regard to comprehension of details and then only on the multiple-choice task.

Several factors may have led to these findings. First, it may be that, despite the careful effort that was made to differentiate the two texts, they were not as different as they were thought to be; thereby obviating any possible effects for text type. Since the interview text was based upon the lecture text, it retained many lexical and grammatical elements of the lecture text, thus making the two texts nearly identical in those regards. In addition, some of the interviewee's responses were lengthy, perhaps making them too similar to the corresponding passages in the lecture text. Moreover, perhaps the interview text did not incorporate enough elements of negotiated discourse to differentiate it from the lecture.

Another explanation for the lack of significant findings with regard to text type may be that content plays a more significant role in determining L2 comprehension performance than does text type. Support for this argument comes from comparing the results of this study with the results of studies such as Markham and Latham

(1987) in which a main effect for text was obtained when content was varied and text type was held constant. The presence of a main effect for text when content is a variable and the lack of a main effect for text when text type is a variable suggests that it is content rather than text type which determines the comprehensibility of a text.

A third explanation for the lack of significant findings with regard to text type may be that the difficulty and/or length of the texts may have obviated any significant effect for text type. With respect to difficulty, the low mean scores, particularly on items testing comprehension of main ideas, indicate that the texts were difficult to comprehend. It may be that the texts were so difficult that the features of negotiated discourse which were predicted to facilitate the comprehension of the interview text remained undetected. Regarding length, it must be noted that the texts used in this study were two to three times longer than the recommended length of two to three minutes. It may be that the texts were so long that any possible effects of differences between them were superseded by the effects of memory. However, Vandergrift (personal communication) points out that while this may be true for the beginning level learners it is not necessarily true for the advanced-intermediate subjects.

A fourth explanation for the lack of significant findings with regard to text type may be that subjects in this study were only allowed to listen to the texts once. In contrast to the present study, subjects in Shohamy and Inbar (1991) were allowed to listen to the

texts twice and a significant effect for text type was obtained. It may be that the factors which are hypothesized to facilitate the comprehension of dialogue texts may become available to L2 listeners only during a second or subsequent listenings.

One final explanation for the lack of significant findings with regard to text type may be the relationship between the languages involved in the L2 learning context under investigation. On the one hand, the L2 language context in Shohamy and Inbar (1991) involved languages which presumably share few, if any, elements--Hebrew and English. In contrast, the L2 learning context in the present study involved languages which share a number of elements, particularly with regard to lexis--English and Spanish. Based on this observation, it could be argued that text type may be a factor in L2 listening comprehension performance when the native and target language are unrelated, but not when there are some similarities between the native and target languages.

With respect to the second research question, the results of this study revealed that assessment task is a significant factor in L2 listening comprehension performance. The results of the study also tended to support the hypothesis that subjects receiving the multiple-choice task would score higher than subjects receiving either the open-ended or cloze task. The most plausible explanation for this, as posited by Shohamy (1984) and Wolf (1991), is that subjects' scored higher on the multiple-choice task because it involved recognition of the correct response whereas both the open-ended and cloze tasks involved retrieval and production of the

correct response. The one exception to this pattern of performance across assessment tasks, concerns the comprehension of details. In this case, subjects receiving the multiple-choice and cloze tasks scored higher than those receiving the open-ended task. A possible explanation for this result may be that the additional context provided by the cloze passage may have helped facilitate reconstruction of the text, which in turn may have facilitated the recall of more details.

Regarding the third research question, the results of the study revealed that target language experience is a significant factor in L2 listening comprehension performance from a quantitative standpoint. Furthermore, the hypothesis that the advanced-intermediate level subjects would score higher than the beginning level subjects was supported in all cases. Despite the significant role played by target langauge experience from a quantitative standpoint, the results revealed that target language experience is not a significant factor in L2 listening comprehension performance from a qualitative standpoint. That is, both levels of subjects exhibited similar patterns of performance across all text types and assessment tasks. The most plausible explanation for these findings is that the advanced-intermediate level subjects' greater experience with and exposure to Spanish allowed them to comprehend more of the texts than the beginning level subjects, but were not a factor in determining patterns of performance.

Finally, with respect to the fourth research question, results of the present study revealed that, with one exception, the three

variables did not interact to a significant degree. This indicates that the effects of each of the three variables were consistent across the other two variables, which in turn indicates uniform patterns of performance. The lone interaction was obtained between text type and assessment task on items testing comprehension of details. Specifically, subjects receiving the multiple-choice task recalled more details after viewing the interview text than they did after the lecture text. No such difference was obtained for subjects receiving either the open-ended or the cloze task.

A possible explanation for this finding may be that items testing comprehension of details are indeed sensitive to differences in text type, but that these differences are revealed only on certain assessment tasks due to the nature of the assessment tasks themselves. In this case, it may be that subjects viewing the interview text were able to demonstrate their superior comprehension of details on the multiple-choice task because it required only recognition of the correct response but were unable to do so on the open-ended and cloze tasks because they required retrieval and production of the correct response.

Despite the lack of significant findings with regard to text type, several recommendations for L2 listening comprehension assessment and pedagogy can be derived from the results of this study. First, given the discrepancies between the findings of this study and the findings of Shohamy and Inbar (1991), much more research is needed in order to determine whether there is in fact a relationship between text type and L2 listening comprehension and whether any

conclusions regarding this relationship are generalizable or are specific to a particular study. This future research might include studies employing text types other than lecture and interview. In addition, the implication that text type might be a factor in comprehension of details but not in the comprehension of main ideas needs to be investigated further. Finally, both the present study and Shohamy and Inbar (1991) investigated the effects of text type from a purely quantitative perspective. It may be that effects for text type might be revealed as the result of a qualitative analysis of summaries or recall protocols.

The second recommendation suggested by the results of this study is that, since assessment task has been shown to be a significant factor in L2 listening comprehension performance, valid L2 listening comprehension assessment instruments should include a variety of different assessment tasks rather than just one type of assessment task. A similar recommendation was made by Wolf (1991) with regard to L2 reading comprehension and appears to apply equally well to L2 listening comprehension. If only one type of assessment task is used, listening comprehension performance may be overestimated in the case of recognition-based assessment tasks or underestimated in the case of production-based assessment tasks. If a combination of assessment tasks is employed, a more balanced and hence, more accurate assessment of L2 listening comprehension performance will result.

Another recommendation suggested by the results of this study is that assessment tasks should test information from different levels

of the text, such as main ideas and details. This recommendation echoes the recommendation of Shohamy and Inbar (1991) that valid listening comprehension tests could include both local and global question types, depending on the purpose of the test. If comprehension of only one type of information is assessed, certain aspects of L2 listening comprehension performance may be obscured. By testing comprehension of the full range of information in the text, a more complete and accurate assessment of L2 listening comprehension performance will result.

A fourth recommendation suggested by the results of this study is that items testing information from different levels of the text should be analyzed separately. This recommendation is based on the fact that in this study, subjects performed differently on items testing comprehension of main ideas and items testing comprehension of details. Such a finding would not have resulted if performance on the two types of items had not been analyzed separately. This suggests that combining items testing different types of information in a single analysis may obscure certain aspects of L2 listening comprehension performance and that by analyzing these items separately, a more complete and accurate assessment of L2 listening comprehension performance will result.

One final recommendation supported, though not suggested, by the results of this study is that comprehension of main ideas and details be assessed after separate listenings. Vandergrift (personal communication) argues that good L2 listening pedagogy should emphasize comprehension of main ideas during the first listening

and comprehension of details during a second and subsequent listenings. This suggestion is supported by the observation made by Shohamy and Inbar (1991) that subjects had a better chance of successfully completing the task if they checked their hypotheses about the overall theme of the passage during the first listening and filled in gaps in the information by paying attention to details during the second listening.

These recommendations and many other issues regarding L2 comprehension assessment need to be addressed more fully. It is hoped that the insights provided by this study and its findings will serve as catalysts for further research into L2 listening comprehension assessment. It is only through expanding our knowledge of L2 listening comprehension assessment that we will be able to develop instruments which give us a more complete and accurate assessment of L2 listening comprehension performance.

Notes

1) This paper was compiled from the author's dissertation (Berne, 1992) and is an extended version of a paper presented at the annual meeting of the American Association of Teachers of Spanish and Portuguese, 11-13 August, 1993, Phoenix, AZ. Earlier versions of this paper were presented at the Third Conference on the Relationship between Second Language Acquisition and Foreign Langauge Learning, 26-28 February, 1993, West Lafayette, IN. and the annual meeting of the American Association for Applied Linguistics, 16-19 April, 1993, Atlanta, GA.

2) Many thanks to Courtney Harrison, Laurens Vandergrift, and the members of the University of North Dakota Faculty Writing Seminar for their valuable comments and insights in the preparation of this paper.

3) For a detailed examination of the stimulus materials and assessment tasks developed for this study, refer to Berne (1992).

References

Bacon, S. M. (1991). The relationship between gender, comprehension, processing strategies, cognitive and affective response in foreign-language listening. Paper presented at the annual meeting of the American Association of Teachers of Spanish and Portuguese, Chicago, IL.

Berne, J. E. (1992). The effects of text type, assessment task, and target language experience on foreign language learners' performance on listening comprehension tests. Unpublished Ph.D. dissertation. University of Illinois at Urbana-Champaign.

Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. TESOL Quarterly, 18, 441-469.

Chaudron, C. (1983). Simplification of input: Topic reinstatements and their effects on L2 learners' recognition and recall. TESOL Quarterly, 17, 437-458.

Herron, C. A. & Seay, I. (1991). The effect of authentic oral texts on students listening comprehension in the foreign language classroom. Foreign Language Annals, 24, 487-495.

Hudson, T. (1982). The effects of induced schemata on the "short circuit" in L2 reading: Non-decoding factors in L2 reading performance. Language Learning, 32, 1-31.

Krashen, S. D. (1982). Principles and practice in second language acquisition. New York: Pergamon Press.

Lee, J. F. (1987). Comprehending the Spanish subjunctive: An information processing perspective. Modern Language Journal, 71, 50-57.

Lee, J. F. (1988). Models for exploring non-native reading. Paper presented at Research Perspectives on Adult Language Learning and Acquisition (RP-ALL), Columbus, OH.

Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass & C. G. Madden (Eds.) Input in second language acquisition. Cambridge, MA: Newbury House.

Markham, P. & Latham, M. (1987). The influence of religion-specific background knowledge on the listening comprehension of adult second-language students. Language Learning, 37, 157-170.

Postovsky, V. A. (1981). The priority of aural comprehension in the language acquisition process. In H. Winitz (Ed.), The comprehension approach to foreign language instruction (pp. 170-186). Rowley, MA: Newbury House.

Rubin, R.B. & Roberts, C. V. (1987). A comparative examination and analysis of three listening tests. Communication Education, 36, 142-153.

Seright, L. (1985). Age and aural comprehension achievement in francophone adults learning English. TESOL Quarterly, 19, 455-473.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. Language Testing, 1, 147-170.

Shohamy, E. & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. Language Testing, 8, 23-40.

Steffensen, M. S. & Joag-Dev, C. (1984). Cultural knowledge and reading. In J. C. Alderson & A. H. Urquart (Eds.), Reading in a foreign language language (pp. 48-61). New York: Longman, Inc.

VanPatten, B. (1990). Attending to form and content in the input. Studies in Second Language Acquisition, 12, 287-301.

Wolf, D. F. (1991). The effects of task, language of assessment, and target language experience on foreign language learners' performance on reading comprehension tests. Unpublished Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

Table 1

ANOVA Summary Table for Overall Scores.

Source of Variation	df	Sum of Squares	Mean Square	F	p
Text Type	1	.129	.129	.049	.8246
Assessment Task	2	65.294	32.647	12.498	.0001*
Target Language Experience	1	147.055	147.055	56.294	.0001*
Text Type X Assessment Task	2	6.624	3.312	1.268	.2843
Text Type X Target Language Experience	1	.127	.127	.049	.8259
Assessment Task X Target Language Experience	2	2.383	1.192	.456	.6345
Text Type X Assessment Task X Target Language Experience	2	2.577	1.288	.493	.6116
Residual	159	415.354	2.612		

* $p < .05$

N = 171

Table 2

Scheffé's Test for Assessment Task based on Overall Scores.

Assessment Tasks	Difference	Critical Difference	p
Multiple-Choice vs. Open-Ended	1.662	.762	.0001*
Multiple-Choice vs. Cloze	.999	.736	.0044*
Open-Ended vs. Cloze	.663	.750	.0954

* $p < .05$

Table 3

Scheffé's Test for Target Language Experience based on Overall Scores.

Target Language Experience	Difference	Critical Difference	p
Beginning Level vs. Advanced-Intermediate Level	1.870	.504	.0001*

* $p < .05$

Table 4

Item-to-Total Correlations Across Entire Data Set.

Pearson Rank Order Correlation Coeffcients

Item	Item-to-Total Correlation	Item	Item-to-Total Correlation
1	.41	6	.46
2	.40	7	.47
3	.44	8	.40
4	.45	9	.47
5	.51	10	.36

Spearman Rank Order Correlation Coefficients

Item	Item-to-Total Correlation	Item	Item-to-Total Correlation
1	.43	6	.44
2	.41	7	.47
3	.42	8	.38
4	.44	9	.47
5	.50	10	.33

Table 5

ANOVA Summary Table for Adjusted Scores.

Source of Variation	df	Sum of Squares	Mean Square	F	p
Text Type	1	.002	.002	.001	.9751
Assessment Task	2	41.732	20.866	9.974	.0001*
Target Language Experience	1	106.765	106.765	51.033	.0001*
Text Type X Assessment Task	2	6.052	3.026	1.446	.2385
Text Type X Target Language Experience	1	1.228	1.228	.587	.4447
Assessment Task X Target Language Experience	2	1.754	.877	.419	.6583
Text Type X Assessment Task X Target Language Experience	2	1.802	.901	.431	.6508
Residual	159	332.643	2.092		

* $p < .05$

N = 171

Table 6

ANOVA Summary Table for Main Idea Scores.

Source of Variation	df	Sum of Squares	Mean Square	F	p
Text Type	1	.067	.067	.075	.7839
Assessment Task	2	30.225	15.113	17.056	.0001*
Target Language Experience	1	51.701	51.701	58.349	.0001*
Text Type X Assessment Task	2	1.492	.746	.842	.4329
Text Type X Target Language Experience	1	.006	.006	.007	.9352
Assessment Task X Target Language Experience	2	.640	.320	.361	.6974
Text Type X Assessment Task X Target Language Experience	2	1.960	.980	1.106	.3334
Residual	159	140.887	.886		

* $p < .05$

N = 171

Table 7

ANOVA Summary Table for Detail Scores.

Source of Variation	df	Sum of Squares	Mean Square	F	p
Text Type	1	.381	.381	.311	.5780
Assessment Task	2	11.277	5.639	4.598	.0114*
Target Language Experience	1	24.367	24.367	19.869	.0001*
Text Type X Assessment Task	2	9.655	4.827	3.936	.0215*
Text Type X Target Language Experience	1	.078	.078	.064	.8010
Assessment Task X Target Language Experience	2	1.386	.693	.565	.5695
Text Type X Assessment Task X Target Language Experience	2	.052	.026	.021	.9791
Residual	159	194.992	1.226		

* $p < .05$

N = 171

Table 8

Contrasts in Mean Scores for Details by Assessment Task.

Contrast	df	Sum of Squares	Mean Square	F	p
Multiple- Choice vs. Open-Ended	1	9.710	9.710	7.917	.0055*
Multiple-Choice vs. Cloze	1	.229	.229	.187	.6659
Open-Ended vs. Cloze	1	7.358	7.358	6.000	.0154*

* $p < .05$

Table 9

Contrasts in Mean Scores for Details by Target Language Experience.

Contrast	df	Sum of Squares	Mean Square	F	p
Beginning Level vs. Adv-Intermediate Level	1	24.367	24.367	19.869	.0001*

* $p < .05$

Table 10

Contrasts in Mean Scores for Details by Text Type and Assessment Task.

Contrast	df	Sum of Squares	Mean Square	F	p
Lecture Multiple-Choice vs. Interview Multiple-Choice	1	8.349	8.349	6.808	.0099*
Lecture Open-Ended vs. Interview Open-Ended	1	.800	.800	.652	.4205
Lecture Cloze vs. Interview Cloze	1	.872	.872	.711	.4004
Lecture Open-Ended and Cloze vs. Interview Open-Ended and Cloze	1	1.665	1.665	1.358	.2457
Lecture Multiple-Choice, Open-Ended, and Cloze vs. Interview Multiple-Choice, Open-Ended, and Cloze	1	.381	.381	.311	.5780

* $p < .05$